Exercise 1: VC Dimension

Consider a binary classification learning problem with feature space $\mathcal{X} = \mathbb{R}$ and label space $\mathcal{Y} = \{-1, 1\}$. Moreover, let

$$\mathcal{H} = \{h_{a,b} : \mathcal{X} \to \mathcal{Y} \,|\, a, b \in \mathbb{R}, a \leq b\}$$

be the hypothesis space of interval classifiers on the reals, where $h_{a,b}(x) = 1$ for $x \in [a, b]$ and $= -1$ otherwise, and

$$\mathcal{H}' = \{h_c : \mathcal{X} \to \mathcal{Y} \,|\, c \in \mathbb{R}\}$$

be the hypothesis space of neighborhood classifiers, where $h_c(x) = 1$ for $x \in [c - 1, c + 1]$ and $= -1$ otherwise.

An assignment for a set of points $x_1, \ldots, x_N \in \mathcal{X}$ by means of $h \in \mathcal{H}$ (or $h \in \mathcal{H}'$) is the vector $(h(x_1), \ldots, h(x_N))^\top \in \mathcal{Y}^N$. A set of points is shattered by $\mathcal{H}$ (or $\mathcal{H}'$) if we can find for any $(y_1, \ldots, y_N)^\top \in \mathcal{Y}^N$ a hypothesis $h \in \mathcal{H}$ (or $h \in \mathcal{H}'$) such that

$$(y_1, \ldots, y_N)^\top = (h(x_1), \ldots, h(x_N))^\top.$$

The maximal number of points N which can be shattered by $\mathcal{H}$ (or $\mathcal{H}'$) is the VC-dimension of $\mathcal{H}$ (or $\mathcal{H}'$) and denoted by $VC_p(\mathcal{H})$ (or $VC_p(\mathcal{H}')$), where p is the dimension of $\mathcal{X}$.

(a) Show that $\mathcal{H}$ is "richer" than $\mathcal{H}'$ in the sense that $\mathcal{H}' \subseteq \mathcal{H}$ but $\mathcal{H} \not\subseteq \mathcal{H}'$.

(b) Consider three (arbitrary) points $x_1, x_2, x_3 \in \mathcal{X}$. How many assignments are in general possible for these three points?

(c) Now assume that the points are such that $x_1 < x_2 < x_3$. Is there an assignment which cannot be generated by some $h \in \mathcal{H}$?

(d) What does the latter mean for the VC-dimension of $\mathcal{H}$ and $\mathcal{H}'$?

(e) Specify two points $x'_1, x'_2 \in \mathcal{X}$ such that they can be shattered by $\mathcal{H}'$.

(f) What does the latter mean for the VC-dimension of $\mathcal{H}'$ and $\mathcal{H}$?